Virtual Signature Testing Capabilities for Ground Targets

February 2004

Teresa Gonda, Dr. Thomas Meitzler, Dr. Dave Bednarz, EJ Sohn , Erik Polsen, Roger Evans
TARDEC
6501 E. Eleven Mile Rd
AMSRD-TAR-R   MS 263
Warren, MI 48397-5000

Randy Wheeler
US Army Materiel Systems Analysis Activity
392 Hopkins Rd
Aberdeen Proving Ground, MD 21005-5071

Dr. Jeffrey S. Sanders
Signature Research, Inc.
150 West Park Loop, Suite 204
Huntsville, AL 35806

ABSTRACT(U)

The Army has an ambitious and aggressive schedule laid out to develop the Future Combat Systems (FCS). Due to extensive testing requirements in all areas, an effort has been made in the last several years to push the capabilities and validation of signature modeling tools that aid in both the design and performance assessment processes. Existing thermal simulations have been leveraged, as have existing synthetic scene generation capabilities, to provide data for testing out signature concepts. In addition, these tools are being looked at and validated for use as a supplement testing capability for contract acceptance tests between the lead system integrator for FCS and the vehicle developer/integrators. This paper will show the tools currently in place and explain the validation tests performed in 2004 and 2005.

1.0 Need for Virtual Testing in FCS

With current operations in Afghanistan and IRAQ, survivability is at the top of everyone's mind. There is a growing urgency to develop a process for determining an optimal recipe for overall soldier and system survivability. The problem with signature management is that it is an exceptionally difficult area to quantify and analyze and hence it is difficult to develop concrete recommendations for vehicle developers to meet. The result is that leaders often trade-off requirements in lieu of more tangible technologies or techniques. This trading away is not done on the basis of a clear understanding of the future threat and a comprehensive survivability plan, but is based more on the inability of decision makers to concretely prove the benefits that signature management provides. This inability is due to a lack of robust measurement tools and techniques that define performance against the threat and balance it against weight, cost, volume, and logistics constraints.

However, intuitively, there is an inherent soundness in trying to become less detectable and the over arching Army requirements documents do reflect this position. What then, is the path forward in trying to understand the role of signature management in survivability? The Army is developing high fidelity physics-driven simulation tools and techniques to supplement and improve the testing or quantifying of IR and visual signatures of systems. It is the hope that this new capability will allow researchers to better understand the phenomenon

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 FEB 2004**	2. REPORT TYPE **Journal Article**	3. DATES COVERED **06-09-2003 to 01-01-2004**

4. TITLE AND SUBTITLE **Virtual Signature Testing Capabilities for Ground Targets**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Teresa Gonda; Thomas Meitzler; Dave Bednarz; EJ Sohn; Erik Polsen**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Signature Research, Inc.,150 West Park Loop,Suite 204,Huntsville,AL,35806**	8. PERFORMING ORGANIZATION REPORT NUMBER **; #14561**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **U.S. Army TARDEC, 6501 East Eleven Mile Rd, Warren, Mi, 48397-5000**	10. SPONSOR/MONITOR'S ACRONYM(S) **TARDEC**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **#14561**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Army has an ambitious and aggressive schedule laid out to develop the Future Combat Systems (FCS). Due to extensive testing requirements in all areas, an effort has been made in the last several years to push the capabilities and validation of signature modeling tools that aid in both the design and performance assessment processes. Existing thermal simulations have been leveraged, as have existing synthetic scene generation capabilities, to provide data for testing out signature concepts. In addition, these tools are being looked at and validated for use as a supplement testing capability for contract acceptance tests between the lead system integrator for FCS and the vehicle developer/integrators. This paper will show the tools currently in place and explain the validation tests performed in 2004 and 2005.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Public Release**	**15**	

around detection and with more robust data, be able to better analyze and determine the resulting vehicle vulnerability. Meanwhile, the Lead System Integrator (LSI) for the FCS program is leveraging this work and has set out a plan to supplement the scheduled field tests of physical prototypes (limited by cost and schedule) with numerous simulations that look at the prototypes under something more representative of the numerous conditions it will be viewed in.

1.1 The Metric

Currently, the most common way of describing a vehicle's signature is to speak in terms of detection by using the Probability of Detection. Lacking validated human observer computer models and simulations to determine these values for the purpose of evaluating a vehicle or treatment, we most often resort to human observer trials. In these trials numerous subjects are measured while they determine whether or not they detect vehicles in the field. They do this for a statistically significant number of runs as the vehicle is placed in different locations in the field of view (or not at all). The results are recorded and a probability of detection can then be determined and broken down by range, vehicle, orientation or any number of conditions. This procedure is a large undertaking and provides a good degree of confidence psychologically, but there are limitations to this technique.

1.2 The Limitations of Field Trials

The first and most obvious problem with field trials is the inability to test concept vehicles that do not physically exist. Prototypes must first be built and this lengthens or cripples the design process and can be expensive. The result is it reduces the ability of a program to optimize designs. Secondly, tests in the field use 50 to 100 observers, are conducted in remote locations, and last several weeks. The tests usually run into a half million dollars to execute and have the data reduced into probabilities of detection. Also, the results are then limited to only that location and are subject to the weather, since the logistics involved in booking a test range and the subjects means that there is very little flexibility if the weather does not cooperate. Further, if the test is classified, then securing appropriate test ranges, security guards and administrators and cleared subjects adds an enormous additional burden to the experiment. Therefore, rain or consistent cloudy weather (if undesirable) often can render the data either useless or can reduce the value greatly.

1.3 The Perception Laboratory

In order to reduce the costs and logistical burdens of field trials, engineers and scientists have moved the tests indoors to perception laboratories. High quality calibrated imagery is taken in the field and later showed to subjects inside a perception laboratory. This approach has many benefits over the field such as experiment repeatability (all subjects see exactly the same images), lower subject logistics burden (more trials in less time, since the vehicle does not have to be moved between each run). The question naturally arises about the validity of this approach. The perception lab experiment has a good track record and represents the trends we see in the field, but rarely the exact result. This is often assumed to be due to the difference in brightness and dynamic range that is in the field, but currently not reproducible in the laboratory. However, if one understands the correlation of a perception lab to the field, the laboratory can be used for the purpose of determining goodness, especially when the experiment is a comparison between two choices, such as a baseline vehicle and one modified. We will discuss more on the perception laboratory and its pedigree later.

Figure 1. Miss Kimberly Lane reviewing test images in the Visual Perception Lab

While the perception laboratory addresses many of the limitations in the field there are still the limitations of location and time (while not as difficult, we are still limited to the amount of data we can collect at a range due mainly to cost). But most importantly, even the perception

laboratory cannot be used for vehicles that do not physically exist…the concept vehicle.

1.4 The Virtual Field Test

Our proposed solution then is the virtual signature test or a simulation-based signature test. Synthetic images of concept vehicles in realistic scenes are generated and then shown to observers in perception laboratories that have tried to quantify their relationship to probabilities in the field.

The Army has been developing a simulation-based signature testing capability that, while still in development, is already proving effective for the Future Combat System (FCS). The Simulation-Based Perception Test (SBPT) program is being developed under the Signature Management for FCS science and technology objective which is now part of the Integrated Survivability Advanced Technology Demonstrator. It is combining two powerful simulations to create physics-driven synthetic scenes for the perception lab.

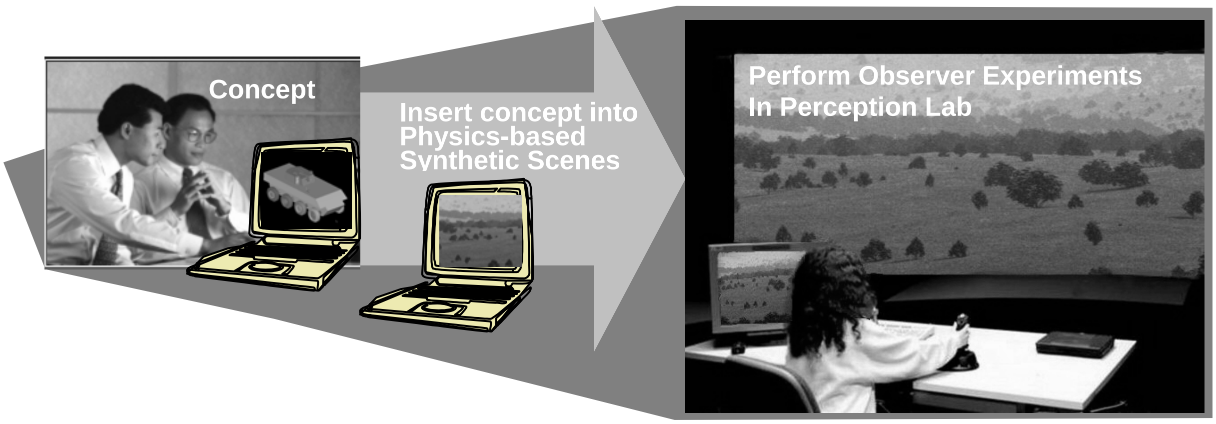


Figure 2. Simulation-based virtual signature test

For this process, the Multi-Service Electro-optic Signature code (MuSES) is used to generate vehicle target signatures for Infrared image which are then inserted into a database for the second simulation software to use. This is CAMEO-SIM, a multi-spectral synthetic scene software developed by INSYS in the U.K. Unlike other high fidelity scene simulators, such as MAYA, CAMEO-SIM is actually a spectral radiance simulation and uses computationally expensive physics to achieve the spectrally accurate photorealism demanded of this application. Since it is multi-spectral, it can display infrared images with the MuSES targets in them plus generate images all the way through visual-including hyperspectral images if desired. These images are then able to be displayed in perception laboratories for controlled environment human observer tests. This simulation-based capability allows for the testing of concept vehicles that do not exist and the ability to view the vehicles under a multiple set of conditions just not practical in the field.

This paper will explore the models used in the simulation, the pedigree of the perception laboratory and the status of addressing the weaknesses and validation of the process.

2.0 Description of Modeling Tools

2.1 PRISM

The Physically Reasonable Infrared Signature Model (PRISM) was one of the top rated and used codes for vehicle thermal signature prediction by 1996. It's final version (3.2) even attempted to become a blend of the two top models by incorporating techniques used in one of the other premier codes, the Georgia Tech Research Institute developed signature code, GTSIG. PRISM was developed largely in FORTRAN and allowed the user the ultimate flexibility in coding up engines and related parts and actually any component given the user could develop a routine for it. Many standard components exist in the library and one can study existing engine

routines and the user's manual to implement new components. A companion software the Faceted Region Editor (FRED) was developed to take much of the burden out of vehicle model development. When PRISM is used properly (as any software tool) it has shown reasonably accurate replication in validation trials of vehicle systems. The final version suffered from multiple improvements over the years however and its user interface dramatically needed an update. In addition, the software validation process began to mature within DoD and it became clear to the Government that with a completely open code, it would be exceptionally difficult to verify and validate individual vehicle system results performed by competitors by anyone other than an expert who was willing to examine every routine in the program for possible tampering. This need to verify contractor results coupled with the need for a dramatic update in computation practices and user interface lead to the decision to develop new software that combined FRED and PRISM.

2.2 MuSES

The goal was to recreate the capability of FRED and PRISM into one program along with a new solver written using the latest in computer science techniques. The new code took the name of the consortium that helped birth it, the Multi-Service Electro Optical Signature (MuSES) code. The new solver was written in C and paid for by the Army and FORD Motor Company under a cooperative research agreement. Further development on the interface and the solver came from the automotive companies' investments, the Air Force, Navy, Army, and from the Small Business Innovative Research program. While it managed many improvements over PRISM over the past several years, those who wished to create dynamic vehicle models with engines, were still encouraged to use PRISM to generate the curves used within MuSES. Only now, with version 7.1 is MuSES able to claim its full independence from PRISM. This latest version allows for user routines to be created and linked to MuSES giving the user the flexibility granted by PRISM and reducing the burden of the verification process for the Government. In development since 1997, MuSES has reached legacy code status, has been and continues to go through verification and validation (V&V) efforts with respect to different tasks, and is now used by the Army on its major vehicle programs. Its ability to link with computational fluid dynamics was recently used to analyze an overheating problem of the mounted mast sight controller box on the Kiowa helicopters in Kuwait.

2.3 CAMEO-SIM

The Camouflage Electro-Optic Simulation System (CAMEO-SIM) was developed by INSYS, Ltd of the United Kingdom for researchers at the Defense Science and Technology Laboratory in Farnborough (dstl). It renders 0.4-20 micron 32-bit physics based synthetic imagery based on a 3D textured geometric representation of the synthetic environment. Unlike commercial renders used in movies or synthetic scene generators for interactive systems such as training, CAMEO-SIM is a first principles simulator working in spectral radiometric space solving the underlying physical equations of radiation transport. This difference is critical for this particular application since it enables the system to become predictive in nature. Inputs to CAMEO-SIM include time of day, weather, material properties (optical and thermophysical), and wavelength(s) to be viewed.[1]

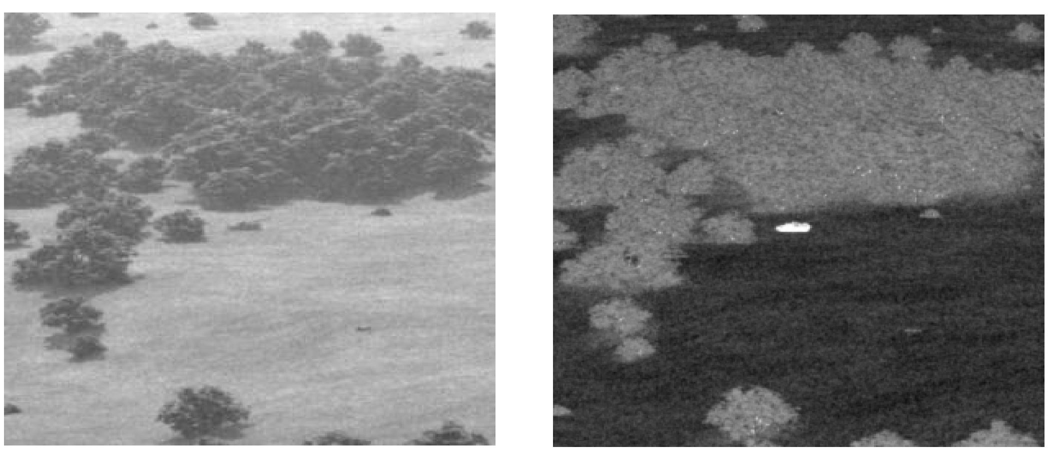


Figure 3. Visual and IR Images from CAMEO-SIM

3.0 Conducting Laboratory Perception Tests.

As mentioned earlier, laboratory tests can be done at a fraction of the cost and in less time and yield results with higher statistical confidence than those done in the field. The laboratory perception tests are performed in a controlled environment, which allows for repeatable experiments and results with a high confidence level.

These tests are conducted using measured imagery collected in the field using film or mega-pixel, high-resolution, digital imagery. Cameras presently available on the market have come very close to equaling the resolution and color depth attainable with film. Six megabyte CCD imaging chips along with the ability to capture imagery in raw 24-bit format, combined with high capacity, portable, storage devices enable high-resolution imagery to be captured at field site locations and easily delivered back to the laboratory. What we are proposing of course is to replace this measured imagery with synthetic imagery at the identical (or even higher) resolution. For now we will speak of the current experience using measured images.

3.1 Displaying the images.

Using high-resolution graphics projectors or monitors, measured imagery is presented in the controlled environment of a laboratory. Repeatability and randomizations offered by the lab environment are not available in a traditional field test. The laboratory randomization of the order of the stimuli removes any potential bias introduced by the order of the presentation of the stimuli in a traditional field test where this type of randomization is not practical.

A photo-simulation test in the lab that mimics a naked eye test is arranged so that the pixel Instantaneous Field-Of-View (IFOV) subtended by monitors (or a projection screen) is less than one minute of arc and the displayed image represents a unity magnification or 1X representation to the subject. Prior to the actual test, the subjects are instructed on the purpose of the test and given a pre-test in which they can become familiar with the imagery and software. None of the pictures used in the pre-test are used in the actual test, however, the images are from the same set. The test protocol is to display an image with a specific time-out, depending on the type of experiment. Often the scene is divided into specific regions and target can appear within one or sometimes more than one of those regions. The subject uses some method (mouse or other device) to identify what he or she thinks is a target, based on the training. The tests are done in a dark room in which the subjects are 'dark-adapted' to maximize contrast differences in the images.

3.2 Calibrating the Display.

In order to get the best match in the laboratory to the field the X and Y chromaticity values of the display must be calibrated to the field. To measure the color chromaticity values in the lab we have used a Photo Research 650 spectrophotometer. The values measured as projected on the monitors are compared to photometric standard values measured in the field. Based on the similarity of photometric measurements between the standards and displayed on two identical monitors, the authors are confident that the color fidelity is accurate. The results of an experiment can be seen in the figures below showing virtual identical X and Y values. The primary physical difference of field versus lab tests is the level of luminance in the lab as compared to the field setting and the dynamic range.

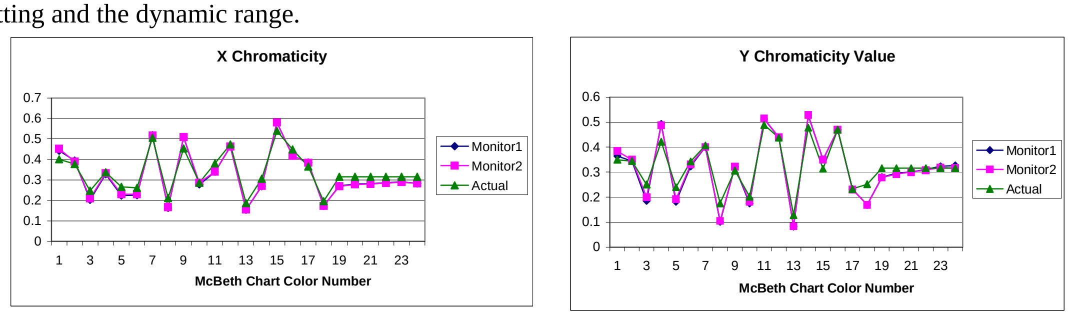


Figure 4 CIE x and y monitor chromaticity calibration charts

Typically when detection tests are done in the laboratory environment, the subjects are dark-adapted and the tests are run under very dim lighting. The use of low light reduces the impact of the effects experienced in the field such as glare and pupil size.

Another benefit to the lab is the ability to resample the imagery to emulate different ranges and the ability to add controlled simulated environmental effects to the images to simulate different degrees of weather variability.

3.1 Designing the Experiment.

When making inferences about differences in a particular factor in a perception experiment, in the laboratory we want to make the experimental error as small possible. This requires that we remove the variability between subjects from the experimental error. The design we use to accomplish this is a factorial experiment run in a randomized complete block. By using this design with the subjects as blocks we form a more homogeneous experimental unit on which to compare different factors. This experimental design improves the accuracy of the comparisons among the different factors by eliminating the variability among the subjects. Within a block, the order in which the treatment combinations are run is randomly determined. In other words, for each subject, the order of the presentation of the imagery is different. It is usually not practical to implement this experimental design in the context of a traditional field test.

3.2 Analyzing the Results.

Below in Table 1 is the ANOVA table for a baseline vehicle and other experimental factors. The power of the experimental design methodology is shown here in that the significance of individual factors and their interactions are available for review. Using this kind of a test, one can obtain not only a math model of the detection probability versus any factor in the test, but, one can also obtain the relative importance of the individual factors and their interactions at various powers.

In Table 1, the first column of the table labeled ‘source’, is the effect or factor(s) in the model, (only first order interactions were considered). The second column shows the type IV sum of squares. The Type VI Sum of Squares factor is used because there are missing cells in our design matrix. The third column, labeled ‘df’, shows the degrees of freedom for each sum of squares. The fourth column labeled ‘Mean Square’, shows the mean square of each effect. This is obtained by dividing the sum of squares for each effect by the degrees of freedom for each effect. The fifth column is the F statistic and shows the F statistic for each effect. The F statistic is obtained by dividing the mean square for each effect by the mean square error term listed at the bottom of the Mean Square column. Column six, labeled ‘Sig’, is the P-value of the F statistic for each effect. The smaller the P-value, the greater the importance of the effect. Table 1 shows that the aspect angle was the least important factor in the experiment and that subject, range, sky condition, and the interaction of range and aspect angle.

Tests of Between-Subjects Effects

Dependent Variable: RANK of RESPONSE

Source	Type IV Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	171682105[a]	79	2173191.202	17.177	.000
Intercept	1277408367	1	1277408367	10096.792	.000
SUBJECT	26976158.0	26	1037544.538	8.201	.000
RANGE	116354292	9	12928254.62	102.187	.000
ASPECT	1125275.323	2	562637.662	4.447	.012
SKY_COND	2522624.471	2	1261312.236	9.970	.000
RANGE * ASPECT	5986618.857	18	332589.936	2.629	.000
RANGE * SKY_COND	5272645.363	18	292924.742	2.315	.001
ASPECT * SKY_COND	2248729.330	4	562182.333	4.444	.001
Error	223301195	1765	126516.258		
Total	1966792305	1845			
Corrected Total	394983300	1844			

a. R Squared = .435 (Adjusted R Squared = .409)

Table 1 ANOVA of test factors

Fig. 8 shows the model generated logistic curve of the probability of detection versus the example baseline vehicle. This curve has the effects of all the various factors 'rolled-up' into it. A logistic curve is the standard psychometric function used to model detection data.

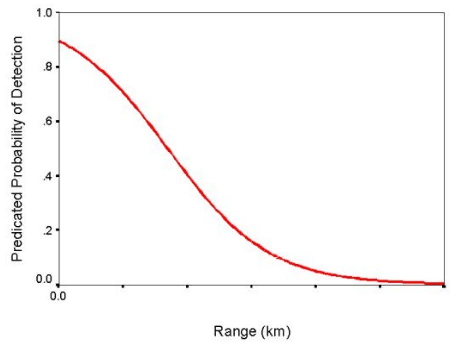


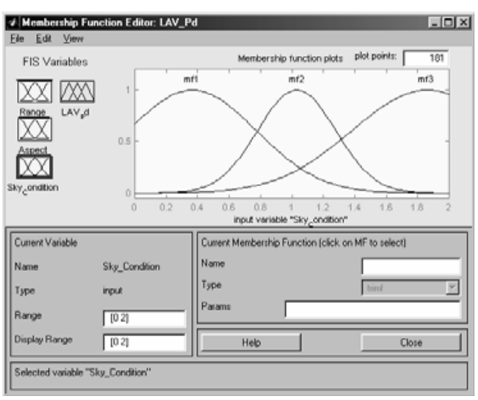
Figure 5: Logistic curve fit to the model from the subject responses

The complete analysis of variance for this experiment is summarized in Table 1. All the main effects except for the aspect angle are significant at the one percent level. The interaction terms are all significant at the one percent level.

3.3 Advanced metrics – Using Fuzzy Logic.

The Fuzzy Logic Approach (FLA) can also be used to model the experimental observer response to the imagery. The FLA and its application to modeling the probability of detection are described in other papers by Dr. Meitzler[2]. The main elements of the model as applied to this sample test are shown below. The correlation obtained in this test was 0.9 between the experimental result and the FLA model predicted value. The 0.9 correlation is between the model built from half the data set and half used as testing.

Figure 6 below shows the variables used in the construction of the 3-input, 1-output Mamdani Fuzzy Logic model. In Figure 7 below the type of membership functions used to simulate the sky condition are shown. When designing the fuzzy logic model the user can select one of several types of membership function. In this case, we chose to use Gaussian bell membership functions.



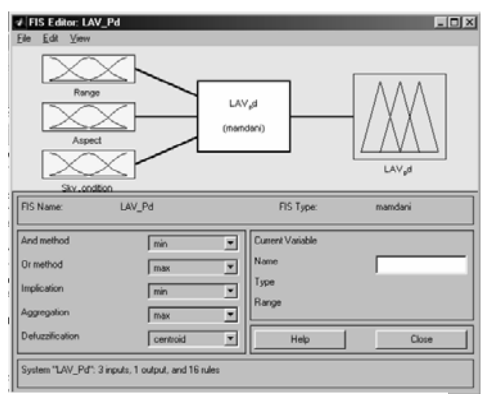


Figure 6 FLA Fuzzy Inference Main Module

Figure 7 FLA membership functions

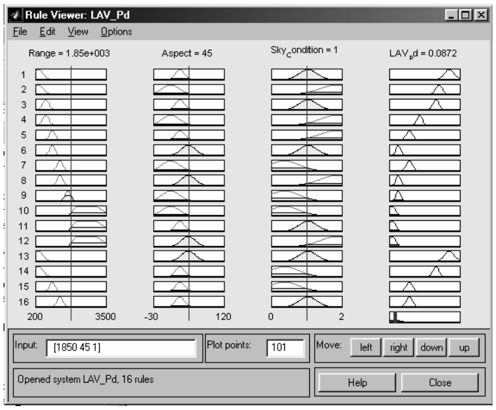


Figure 8: FLA firing diagram

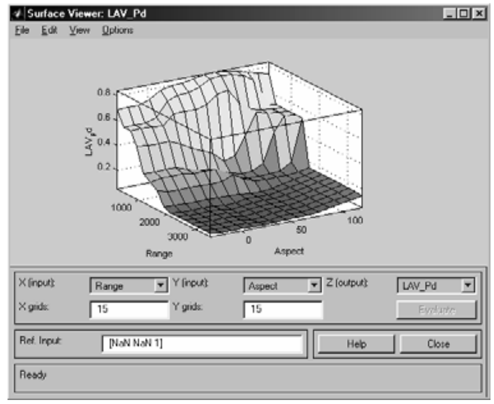


Figure 9: Model surface of Probability of detection versus range and aspect angle

Once the membership function properties and the rule firing strengths have been coded into the model, the program then computes the firing strengths for the various rules and then sums up the results using the centroid method. The rule firings are shown above in Figure 8. The final surface of the fuzzy logic predicted probability of detection versus range and aspect angle is shown in Figure 8

In summary, an advantage of using the photosimulation lab environment is that experimenters are able to archive scenes used in the simulation, so that, at a later time it is possible to rerun the same image data set on a different subject pool. The new subjects may have different training and the images may also be modified by either magnification or adding atmospheric conditions. This provides tremendous cost savings since there is no need to pay for another field test or data collection.

4.0 Validation

4.1 Validation tests for MuSES

MuSES has had and continues to go through extensive verification and validation experiments. Version 6.0 went through a rigorous set of tests and was documented in reported in 2000.[3] One of the plots from the official report is shown in Figure 10. It shows several days of comparison between measured and simulated values of the surface of a thin plate during experiments to establish the most accurate wind model for MuSES. Thin plates are traditionally difficult to model, and the reasoning goes that if you can match the thin plates and compare them against textbook calculations, as well as measured data, then you have achieved an acceptable level of

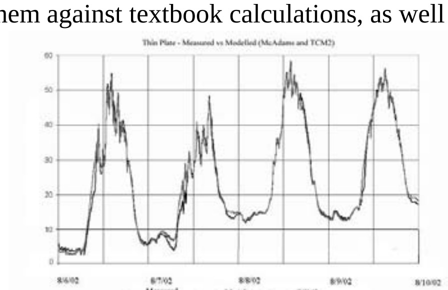


Figure 10 Validation plots during wind model experiment

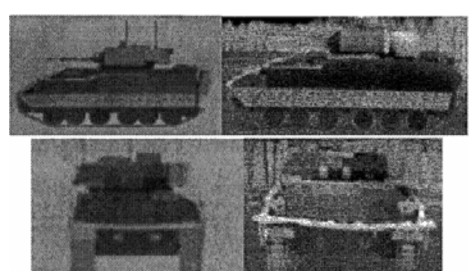


Figure 11 Modeled (left) versus Measured (Right)

goodness. In this case, the TCM2 method of wind calculation (several available in MuSES) has matched up exceptionally well with the measured data and the McAdams method was not that different.
Depending on what is being modeled and the amount of data available to simulate it (surface properties of paint for instance), this type of accuracy is achievable. Simulations such as this are highly dependent on both how something is modeled and the goodness of the input data. However, after successive V&V efforts, what is clear is that the software itself is very sound and gives valid results for the problems it has been asked to address. In addition, each vehicle model must be put through its own V&V process in order to determine its value. Figure 11 shows images rendered of the M2 using the previous legacy code PRISM.[4] MuSES uses similar methods for calculating temperature and has been shown to match PRISM results or better on flat plate tests.

4.2 AMSAA validation of MuSES

Currently, AMSAA has embarked on an extensive testing process to accredit MuSES for use in populating the AMSAA signatures database. The new metric is the RSS Delta T and if AMSAA can accredit MuSES for the purpose of generating the value, it can create an endless amount of data under various conditions to be used in operation analysis such as CASTFOREM.

A robust and complete database of target (US and OPFOR) ΔTRSS values at multiple aspects about the target is required to properly support and feed the Acquire and Acquire-LC methodologies in Army models and simulations (M&S) and predictive IR signature modeling may be the only feasible approach to address this requirement.

AMSAA's approach is to determine the point of diminishing returns in the IR signature model validation process and the cost/benefit ratio of different types of validation data. It is always desirable to have extensive measured signature data to validate predictive signatures against, but collecting the desired data can be prohibitively expensive. However, without some sort of validation data the end user cannot have as high a confidence in the outputs of the predictive model. The validation process for the models in this effort was iterative in the sense that the models were refined against a gradually supplied set of validation data. While extensive validation data was initially collected in field tests, the IR signature modelers were only provided small subsets of the validation data at a time. The purpose of this was to determine the types of validation data that provide the greatest improvements in signature model fidelity and to quantify the incremental signature model improvements.

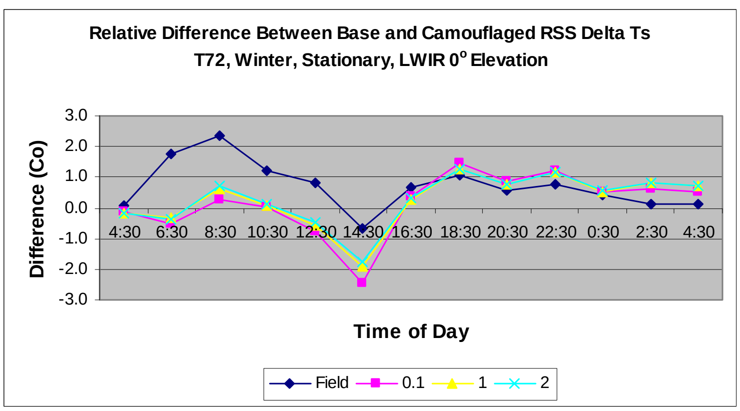


Figure 12 Relative Difference Between the Baseline and Camouflaged RSS delta t Values.

4.3 CameoSim Validation

The United Kingdom Ministry of Defence at the Defence Science and Technology Laboratory (dstl [sic]) has been verifying and validating CAMEO-SIM for several years[5]. Firstly, the code itself is built on an architecture that allows it to generate numerous verification tests to assure the developers and the proponent that the code is producing the expected results. While validation of simulation software has proven complicated, dstl pioneered the development of a strategy that has great promise. This process is built on the premis well known to those in the simulation world that "no model is completely valid", but instead we must focus on validating a simulation for a specific purpose. Since imagery can be rendered at different levels of quality, depending on application, defining metrics of goodness for the application are essential. These metrics can then be used to test the fitness of the simulation results.

dstl developed the Fidelity Investigations and Reporting Environment or FIRE for just this purpose. FIRE is comprised of three groups of metrics meant to tell us something about the images with respect to image quality and its relationship to human perception. The three groups are wavelets metrics, higher order statistics, and a human vision model.

The higher order statistics depend on an image's phase spectrum and encapsulate information about the shape and relative positions of features in the scene. The human vision model, is not all encompassing as that is not feasible at this time, but it does incorporate a popular approach to model[6] how the human visual system can discriminate small changes in the blur of natural scenes; the presence of low contrast test targets in natural scenes; and small changes in the spatial organization of the objects in photographs of natural scenes.

The final group of metrics is extracted by a multi-scale analysis of the image based on wavelet decomposition. Described in "Assessment of synthetic image fidelity "by Gilmore et al "[Error! Bookmark not defined.]

"The parameters deemed important to match between real and synthetic textures for some applications are defined as

1. A measure of overall clutter strength, pc. Similar to average edge strength, but emphasizes the strongest edges and is scale normalized. It is the strongest edges that affect the difficulty of object recognition in clutter.

2. A measure of image smoothness (spatial correlation) called the self-similarity parameter, k. This measure describes how edge strength varies with scale (the size and smoothness of the edge). The dominant edges in an image (large k) have larger scales, whereas rough (uncorrelated) images have dominant small-scale structure.

3. A measure of overall clutter density, a. This is related to the average number of edges in an image, but is scale normalised. This metric tells us how many edges there are in an image, rather than their average strength, pc.

4. A measure of clutter uniformity, d. This measure tells us whether edges are distributed uniformly within the image or whether there are some regions which are more densely populated than others. A high d value indicates a very uniform image whereas a low d value shows a lot of clustering"

dstl analysts have used the metrics in FIRE to analyze synthetic imagery rendered at different levels of fidelity and they have used it to compare synthetic imagery with real-world imagery as seen in Figure 13 and Figure 14. Additionally, some of the metrics in FIRE can be used to compare two spatially correlated images, while others can be used to assess particular characteristics of the image such as clutter level

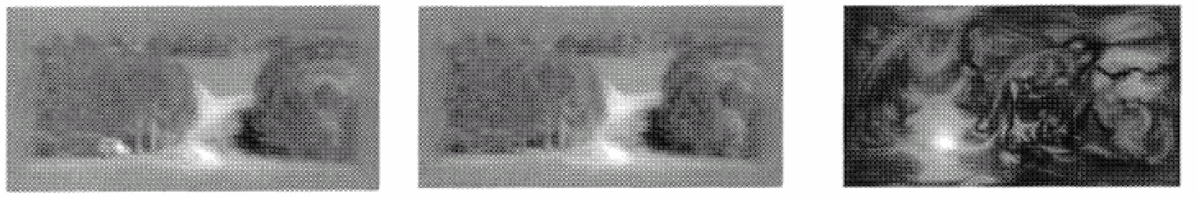


Figure 13 Real Images and Discrimination Map on Right

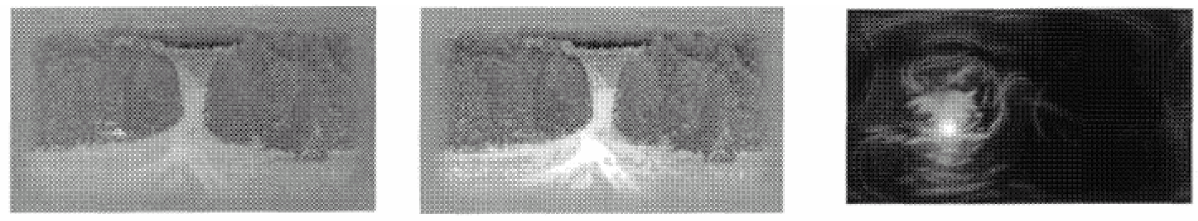


Figure 14 Synthetic Images and Discrimination Map on Right

4.4 Perception Lab Validation

Perception lab experiments have been compared to field experiments for many years. Ashforth and Collins report in 1991 that at that time, they and several others had obtained ratios between 1.5-2.0 in comparing simulated tests using slides and slide projectors versus field experiments[7]. Current projectors and techniques show promise of improving on those figures. What's more, continued improvements in high definition displays mean that the results of the comparisons should continue to improve as well. However, much more work needs to be done to better understand the different techniques used in perception labs and their strengths and weaknesses. While good comparison data is often unavailable due to country restrictions or classification, we do have some to present here.

A 'lab validation' or comparison between perception lab results and field results was performed on a selected set of images extracted from an optical filter test done in collaboration with the Army Material System Analysis Activity (AMSAA) and the Marine Corps at 29 Palms. Figure 15 below is a representative image from that data set.

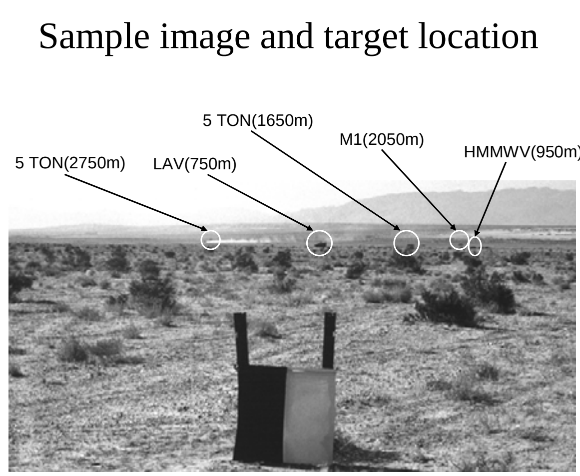


Figure 15 Sample image from data set

For the data set used, slides were the photographic format available and the high resolution graphics projectors currently in use at the TARDEC facility were not installed at that time, but as Figure 16 shows, subjects viewed the images as projected on to a rear-projection screen. The screen to observer distance was chosen so that the angular subtense of the vehicles in the scenes as seen by the observer was the same as seen in the field by subjects. This is a standard rule for all perception tests, in order to compare like-with-like, the geometry of the test setup, as well as image size and magnification, is adjusted so as to present realistic conditions to the observer.

Test setup

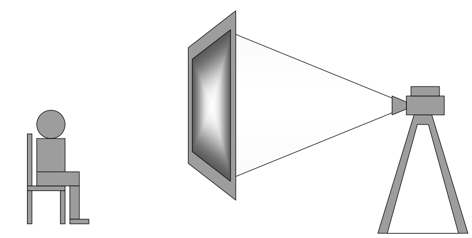


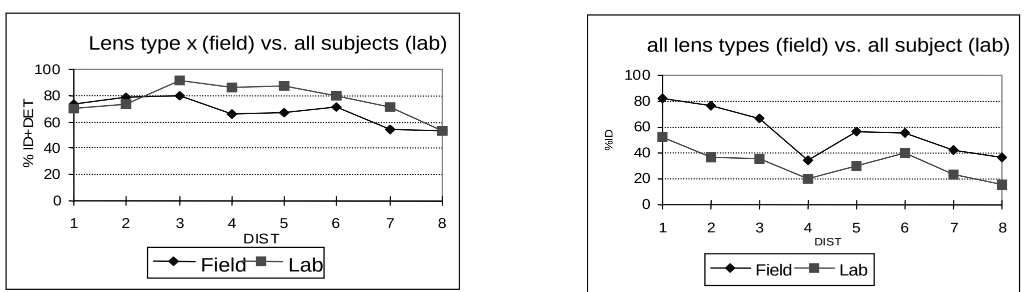
Figure 16 Test geometry used for validation experiment

Figure 17 and Figure 18 below show the comparisons of field results to lab results. The graphs show a generally high correlation, sometimes approaching 0.9 and differing by as little as 10%. More tests have to be done over various data sets and simulated images.

Detection rate over distance

%Identification + % Detection over the distance aggregated over all subject responses

One vehicle identification rate over the distance



Correlation values
From dist-1 through dist-8: 0.58
From dist-3 through dist-8: 0.84

Correlation value: 0.9

Figure 17 Plots of identification rates – field vs. lab

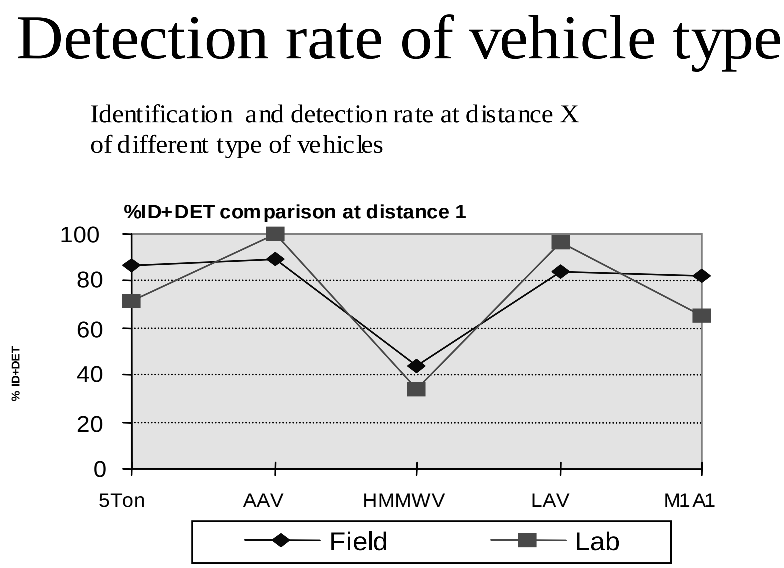


Figure 18 ID and detection rates – field vs. lab

4.4 Joint Field Trial Validation Experiment

The diagram below shows the process for an experiment that will help validate the entire virtual perception test capability by comparing the end metric of Probability of Detection between field, laboratory photosimulation and laboratory tests using synthetic scenes.

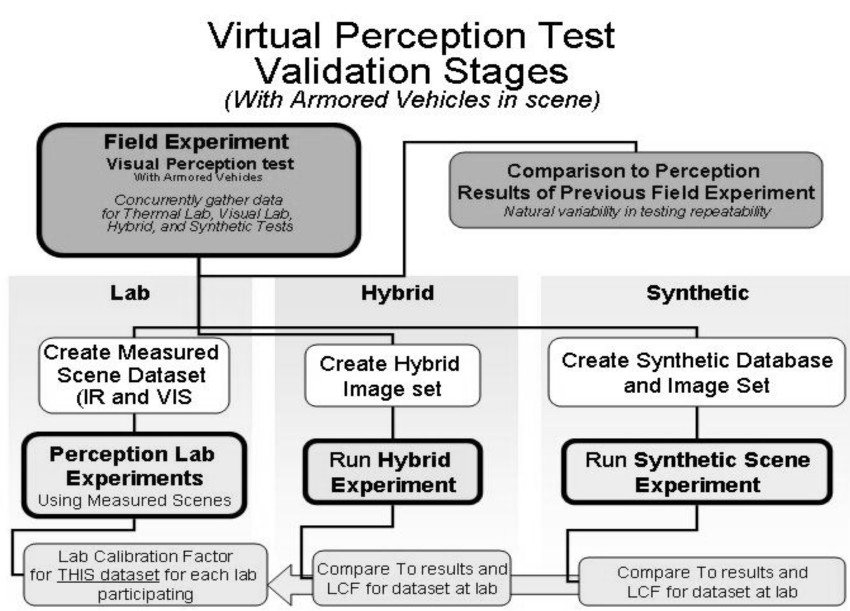


Figure 19 Flow diagram of the Grand validation experiment

The image below shows the landscape of the Eglin test site where a field trial using observers and several vehicles was held. This test will next be recreated in the lab using the Photosimulation process (displaying images identical to those used in the live experiment). Finally, we will recreate the scene in CAMEO-SIM and re-run the experiment in the laboratory with those images.

Figure 20. Eglin observer test image with calibration targets in image.

At the end of this validation exercise we intend to learn much about the entire process of detection as well as how the laboratory tests correlate to the field and of course the strengths and weaknesses of using synthetic imagery for this purpose.

5.0 Summary

We have explained in some detail, the components that are going into the virtual or simulation-based signature testing capability. The results of the validation experiment will be published as they occur and the community will have access to this information. The individual components of the procedure have good levels of validation in and of themselves and the process promises to provide some measure of value even at the onset. With this capability, we will now be able to “test” concept designs while they are still in the computer and prove the merit of competing signature techniques to decision makers.

6.0 References

[1] D. Filbee, A. Kirk, C. Stroud, G. Hutchings, T. Ward, D. Brunnen, "Modeling of High Fidelity Synthetic Imagery for Defence Applications" Proceedings SPIE Conference 4718, 2002 Targets and Backgrounds VIII: Characterization and Representation.

[2] Meitzler, T., Kistner, R., Sohn, E., Bryk, D., and Bednarz, D., "Computing the probability of target detection in dynamic visual scenes containing clutter using the fuzzy logic approach," Opt. Eng., Vol. 37 (7), pp. 1951–1959, 1998.

Meitzler, T., Sohn, E., Singh, H., Elgarhi, A., and Nam, D., Predicting search time in visually cluttered scenes using the fuzzy logic approach," Opt. Eng., 40(9), Sept. 2001, pp. 1844-1851.

Meitzler, T., "Modern approaches to the computation of the probability of target detection in cluttered environments," Ph.D. Dissertation, UM Printing Services, 1996.

[3] T. Gonda, J. Perez, P. Rogers, G. BoBo, R. Evans, J. Jones, "VV&A Process for Muses", Eleventh Annual Ground Target Modeling and Validation Conference, August 2000.

[4] G. Plude, T.Gonda, P. Rynes, J. Jones, "Validation of the M2A2 PRISM Model", 5th Annual Conference on Modeling and Validation of Ground Vehicle Signatures, August 1994.

[5] M.A. Gilmore et al, "CAMEO-SIM: a broadband scene generation system that is fit for purpose", Proceedings Aerosense SPIE Conference Vol. 3699, pp 217-228, Targets and Backgrounds: Characterization and Representation, Orlando FL, April 1999.

D. Filbee, C. Stroud, et al, 'Use of metrics to quantify the accuracy of synthetic images', Proceedings 11th Ground Target Modeling and Validation Conference. Pages 269-279, Houghton, MI 2000.

[6] Rohally, AM., Ahumada, A.J. & Watson, A.B. 'Object detection in natural backgrounds predicted by discrimination performance and models'. Vision Research. 37, pages 3225-3235. 1997.

Watson, A.B. 'Visual detection of spatial contrast patterns: Evaluation of five simple models'. Opt. Express 6, 33. 2000.

[7] M. Ashforth, J.H. Collins, "Determination of Detection Range by Analysis of Recorded Imagery", Stores and Clothing Research Establishment, Flagstaff Rd, Colchester, Essex, Technical memorandum SCRDE/91/6, NATO AC/243/CCD/WG(D), June 1991